fw

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/006,984	12/04/2001	Jan K. Skoglund	020184-000510US	1945

20350 7590 05/13/2005

TOWNSEND AND TOWNSEND AND CREW, LLP
TWO EMBARCADERO CENTER
EIGHTH FLOOR
SAN FRANCISCO, CA 94111-3834

EXAMINER	
ALBERTALLI, BRIAN LOUIS	
ART UNIT	**PAPER NUMBER**
2655	

DATE MAILED: 05/13/2005

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No. 10/006,984	Applicant(s) SKOGLUND ET AL.
	Examiner Brian L Albertalli	Art Unit 2655

-- *The MAILING DATE of this communication appears on the cover sheet with the correspondence address* --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *04 January 2005*.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-21* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☒ Claim(s) *19-21* is/are allowed.

6)☒ Claim(s) *1-4,6,7,10-16 and 18* is/are rejected.

7)☒ Claim(s) *5,8,9 and 17* is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☒ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: _____.

DETAILED ACTION

Response to Amendment

1. The amendments to the claims have been entered. Claims 1, 3, 5, 8-11, 14, 17, and 19-21 have been amended.

Response to Arguments

2. Applicant's arguments filed January 4, 2005, with regard to independent claims 1 and 14 have been fully considered but they are not persuasive.

In the applicant's interview of January 4, 2005 and in the applicant's written response, the applicant has asserted that the term "features" as used in the application has a specific meaning in the art of classifying speech activity using probability models. In the specification, the "features" are described as the signal power in N bands, which are calculated by adding the logarithms of the absolute values of fast Fourier transform (FFT) coefficients derived from the input signal, and normalizing them with the length of the band (page 7, paragraph 35). However, the specification does not limit the definition of "features" to only this definition. As described in the specification (page 5, lines 15-16) a feature is "derived from any tangible characteristic of a digitally sampled signal". Therefore, the term "feature", as interpreted by the examiner herein, encompasses any measurement or calculation that provides information as to the properties of the input signal.

The newly amended independent claims 1 and 14, however, further limit what can possibly be used as "features" extracted from the input signal. Specifically, the

features extracted "cannot alone recreated the digitized signal". Additionally, the newly amended claims require the digitized signal is a digital representation of the signal, but this is clearly anticipated by Sohn et al. (frames of L *samples* of the input signal are used, which are necessarily a digital representation of the signal).

Regarding the limitation that the features extracted from the input signal cannot alone recreated the digitized signal, the examiner agrees that the relied upon portion of Sohn teaches features that could recreate the digitized signal (the features are the true power spectra of the input signal, which is inherently reversible to recreate the input signal). However, Sohn further teaches that when the disclosed voice activity detector is used in a Linear Prediction Coefficient (LPC) based coder, the LPC coefficients can be used to eliminate the computationally expensive DFT operations (page 368, 2^{nd} column, lines 5-8). LP coefficients alone cannot be used to recreate the digitized signal from which they were derived.

Regarding the arguments presented regarding the use of official notice in the previous action, as explained in the previous rejections of claims 1 and 14, Sohn et al. teaches all of the features of the claims, except adapting the first PDF (the probability density function of active speech) based, at least in part on, the plurality of features. A PDF that models active speech is derived (equation 5, probability of noisy speech X given speech is present H1), but Sohn et al. does not teach that the probability that speech is present (equation 5) is adapted over time.

In response to the request that documentary evidence be presented to support the use of official notice, Levinson (*Statistical Modeling and Classification*) is presented.

As taught by Levinson, any features extracted from a speech signal (cepstral features are taught) can be modeled as a random process whose statistical properties can be estimated (page 1, last paragraph). In any class based decision model (corresponding to the active speech class and the inactive speech class presented by the applicant), a statistical decision can be made to what class input features belong to (page 2, 2nd paragraph, lines 4-5). Levinson further discloses that the performance of a speech processing algorithm depends critically on the accuracy of the class conditional density functions, and that the more data that is used to estimate these density functions, the more accurate the classifications will be (page 3, section 11.2.4, lines 3-9).

Therefore, by modifying Sohn et al. to actively adapt the first PDF model for active speech, more data would be collected to estimate the class conditional density function (the probability H1 that the input was active speech), which would increase the probability of correct speech/nonspeech decisions.

For the reasons given above, the rejections to claims 1 and 14 as being unpatentable in view of Sohn et al. are upheld.

Applicant's arguments regarding claims 4 and 16 have been fully considered but they are not persuasive. Although Sohn et al. teaches a mean square approach in the adapting step, this causes the noise model to converge towards the actual noise, page 368, first column, lines 2-4). This would necessarily *increase a likelihood*. That is, the likelihood that a noise frame was correctly identified as noise would increase as the noise model converged to the actual noise.

Therefore, the rejection to claims 4 and 16 are upheld.

3. Applicant's arguments, see 35 U.S.C. 103 Rejection, Sohn et al. in view of Huang et al. section, and see 35 U.S.C. 103 Rejection, Sohn et al. in view of Paez et al. section, filed January 4, 2005, with respect to claims 8, 9, 17, and 19-21 have been fully considered and are persuasive. The rejections of claims 8, 9, 17, and 19-21 have been withdrawn.

Furthermore, the arguments with respect to claim 5 (page 11, last line to page 12, line 2) are persuasive, therefore the rejection of claim 5 is withdrawn.

Specification

4. The amendments to the specification overcome the rejections made in the previous office action. The objections to the specification are withdrawn.

Claim Rejections - 35 USC § 112

5. The amendment to claim 5 overcomes the rejection under 35 U.S.C. 112, 2nd paragraph made in the previous office action. The rejection to claim 5 under 35 U.S.C. 112, 2nd paragraph is withdrawn.

Claim Rejections - 35 USC § 103

6. The text of those sections of Title 35, U.S. Code not included in this action can be found in a prior Office action.

Claims 1-4, 6-7, 10-11, 13-16, and 18 are rejected under 35 U.S.C. 103(a) as being unpatentable over Sohn et al. (*A Voice Activity Detector Employing Soft Decision Based Noise Spectrum Adaptation*), in view of Levinson (*Statistical Modeling and Classification*).

In regard to claim 1, Sohn et al. discloses a method for detecting speech activity for a signal, the method comprising the steps of:

extracting a plurality of features from a digitized signal (LP coefficients, page 368, 2^{nd} column, lines 5-8), wherein:

the plurality of features alone cannot recreate the digitized signal (LP coefficients alone cannot be used to recreate the digitized signal from which they were derived), and

the digitized signal is a digital representation of the signal (frames of L *samples* of the input signal are used, page 365, section 2, lines 12-15);

modeling a first and a second probability density functions (PDFs) of the plurality of features, wherein:

the first PDF models active speech features for the digitized signal (equation 5, probability of noisy speech X given speech is present H1), and

the second PDF models inactive speech features for the digitized signal (equation 4, probability of a noisy speech X, given speech is absent H0);

adapting the second PDF to respond to changes in the digitized signal over time (the noise spectrum is continuously updated, equation 16 and page 368, first column, lines 6-8);

probability-based classifying of the digitized signal based, at least in part, on the plurality of features (decision rule is used to differentiate between silence and noise, equation 7); and

distinguishing speech in the digitized signal based, at least in part, upon the probability-based classifying step (the decision rule is a decision whether speech is present H1, or absent H0, see page 366, first column, lines 11-12).

Sohn et al. does not disclose that the speech PDF is adapted.

Levinson discloses any features extracted from a speech signal (cepstral features are taught) can be modeled as a random process whose statistical properties can be estimated (page 1, last paragraph). In any class based decision model (corresponding to the active speech class and the inactive speech class presented by the applicant), a statistical decision can be made to what class input features belong to (page 2, 2nd paragraph, lines 4-5). Levinson further discloses that the performance of a speech processing algorithm depends critically on the accuracy of the class conditional density functions, and that the more data that is used to estimate these density functions, the more accurate the classifications will be (page 3, section 11.2.4, lines 3-9).

Therefore, by modifying Sohn et al. to actively adapt the first PDF model for active speech, more data would be collected to estimate the class conditional density function (the probability H1 that the input was active speech), which would increase the probability of correct speech/nonspeech decisions.

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Sohn et al. to adapt the speech PDF in addition to the noise PDF, since the statistics of the speech would change over time. This would make the speech PDF model the actual speech more accurately, which would increase the probability of correct speech/non-speech decisions.

In regard to claim 2, Sohn et al. discloses the probability based classifying step uses first and second PDFs (equation 7, classification decision is dependent on PDFs given in equations 4 and 5).

In regard to claim 3, Sohn et al. discloses the modeling step comprises a step of determining a mathematical model (PDFs) for the digitized signal from the plurality of features (the variances of the noise and speech are determined from the power spectra of the noise, equations 1 and 2; which are used to determine the PDFs for the signals, page 365, second column, section 2 line 15 through page 366, equation 2).

In regard to claim 4, Sohn et al. discloses the adapting step comprises increasing a likelihood (equation 16, the noise model converges towards the actual noise, page 368, first column, lines 2-4).

In regard to claim 6, Sohn et al. discloses the probability-based classifying step comprises a step of classifying based on likelihood ratio detection (a log likelihood ratio is used for the decision rule, page 366, equation 7).

In regard to claim 7, Sohn et al. discloses the probability-based classifying step comprises applying a log-likelihood ratio test to one of the plurality of features (page 366, equation 7, the log likelihood ratio is based on the variances of the speech and noise, which are determined from the coefficients from the DFT, page 365, second column, section 2 line 15 through page 366, equation 2).

In regard to claim 10, Sohn et al. discloses at least one of the first and second PDFs comprises a plurality of basic density models (page 366, equations 4 and 5, each PDF is the product of L basic density models).

In regard to claim 11, Sohn et al. discloses at least one of the plurality of features is related to power in a spectral band of the signal (DFT coefficients are determined, the coefficients denote the true power spectra of the noise and speech, page 366, first column, line 3).

In regard to claims 13 and 18, Sohn et al. does not explicitly disclose a computer-readable medium having computer-executable instructions for performing the computer-implementable method for detecting speech activity for the signal of claim 1 or 14.

Official notice is taken that it is notoriously well recognized to implement a signal processing method on a computer and to store instructions for implementing the method on a computer readable medium.

It would have been obvious to one of ordinary skill in the art at the time of invention to store the method as disclosed by Sohn et al. on as computer readable code on a computer readable medium, so the method could be implemented on computer.

In regard to claim 14, Sohn et al. discloses a method for detecting sound activity for a signal, the method comprising the steps of:

extracting a plurality of features from a digitized signal (DFT coefficients, page 365, second column, section 2, lines 15-22), wherein:

the plurality of features alone cannot recreate the digitized signal (LP coefficients alone cannot be used to recreate the digitized signal from which they were derived), and

the digitized signal is a digital representation of the signal (frames of L *samples* of the input signal are used, page 365, section 2, lines 12-15);

modeling an active sound probability density function (PDF) of the plurality of features (equation 5, probability of noisy speech X given speech is present H1);

modeling an inactive sound PDF of the plurality of features (equation 4, probability of a noisy speech X, given speech is absent H0);

adapting the inactive sound PDFs to respond to changes in the digitized signal over time (the noise spectrum is continuously updated, equation 16 and page 368, first column, lines 6-8);

probability-based classifying of the digitized signal based, at least in part, on the plurality of features (decision rule is used to differentiate between silence and noise, equation 7); and

distinguishing sound in the digitized signal based, at least in part, upon the probability-based classifying step (the decision rule is a decision whether speech is present H1, or absent H0, see page 366, first column, lines 11-12).

Levinson discloses any features extracted from a speech signal (cepstral features are taught) can be modeled as a random process whose statistical properties can be estimated (page 1, last paragraph). In any class based decision model (corresponding to the active speech class and the inactive speech class presented by the applicant), a statistical decision can be made to what class input features belong to (page 2, 2nd paragraph, lines 4-5). Levinson further discloses that the performance of a speech processing algorithm depends critically on the accuracy of the class conditional density functions, and that the more data that is used to estimate these density functions, the more accurate the classifications will be (page 3, section 11.2.4, lines 3-9).

Therefore, by modifying Sohn et al. to actively adapt the first PDF model for active speech, more data would be collected to estimate the class conditional density function (the probability H1 that the input was active speech), which would increase the probability of correct speech/nonspeech decisions.

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Sohn et al. to adapt the speech PDF in addition to the noise PDF,

since the statistics of the speech would change over time. This would make the speech PDF model the actual speech more accurately, which would increase the probability of correct speech/non-speech decisions.

In regard to claim 15, Sohn et al. discloses the probability-based classifying step uses the active and inactive speech PDFs (equation 7, classification decision is dependent on PDFs given in equations 4 and 5).

In regard to claim 16, Sohn et al. discloses the adapting step comprises a step of increasing a likelihood (equation 16, the noise model converges towards the actual noise, page 368, first column, lines 2-4).

7. Claim 12 is rejected under 35 U.S.C. 103(a) as being unpatentable over Sohn et al. (*A Voice Activity Detector Employing Soft Decision Based Noise Spectrum Adaptation*), hereinafter referred to as Sohn 1, in view of Levinson, as applied to claim 1, above, and further in view of Sohn et al. (*A Statistical Model-Based Voice Activity Detection*), hereinafter referred to as Sohn 2.

Neither Sohn 1 nor Levinson disclose a step of smoothing an activity decision for hangover periods to produce a smoothed activity decision.

Sohn 2 discloses a step of smoothing an activity decision for hangover periods to produce a smoothed activity decision (a smoothing factor obtained by equation 11 is

used to modify the final decision statistic, page 2, second column, paragraphs three and four).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Sohn 1 to smooth the activity decision for hangover periods, in order to prevent the clipping of weak speech tails, as disclosed by Sohn 2 (page 2, first column, section III, lines 1-3).

Allowable Subject Matter

8. Claims 5, 8, 9, and 17 are objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

The following is a statement of reasons for the indication of allowable subject matter:

Regarding claim 5, the examiner agrees that while Sohn et al. teaches the adaptation using previous frames, there is no indication that would suggest to one of ordinary skill in the art at the time of invention to identifying extreme values (high or low) in the previous frames.

Regarding claims 8, 9, and 17, the examiner agrees that since Sohn et al. specifically makes the assumption that the speech and noise signals are Gaussian random processes, there would be no suggestion to one of to one of ordinary skill in the

art at the time of invention to modify Sohn et al. to use either Gaussian mixture models or non-Gaussian models.

9. Claims 19-21 are allowed.

The following is an examiner's statement of reasons for allowance:

The examiner agrees that since Sohn et al. specifically makes the assumption that the speech and noise signals are Gaussian random processes, there would be no suggestion to one of to one of ordinary skill in the art at the time of invention to modify Sohn et al. to use non-Gaussian models

Any comments considered necessary by applicant must be submitted no later than the payment of the issue fee and, to avoid processing delays, should preferably accompany the issue fee. Such submissions should be clearly labeled "Comments on Statement of Reasons for Allowance."

Conclusion

10. **THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any

extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

11. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Brian L Albertalli whose telephone number is (571) 272-7616. The examiner can normally be reached on Mon - Fri, 8:00 AM - 5:30 PM, every second Fri off.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, David Ometz can be reached on (571) 272-7593. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

BLA 5/10/05

DAVID L. OMETZ
PRIMARY EXAMINER